DOCUMENT RESUME

ED 036 185 24 EM 007 771

AUTHOR NITKO, ANTHONY J.
TITLE MEASUREMENT OF INSTRUCTIONAL OUTCOME VS. MEASUREMENT FOR INSTRUCTION: A VIEW OF IPI TESTING PROCEDURES.
INSTITUTION RESEARCH FOR BETTER SCHOOLS, INC., PHILADELPHIA, PA.
SPONS AGENCY OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
BUREAU NO BR-6-2867
PUB DATE 12 SEP 68
CONTRACT OEC-1-7-062867-3053
NOTE 13P.

EDRS PRICE EDRS PRICE MF-$0.25 HC-$0.75
DESCRIPTORS EDUCATIONAL DIAGNOSIS, *INDIVIDUALIZED INSTRUCTION, MEASUREMENT INSTRUMENTS, *MEASUREMENT TECHNIQUES, *PROGRAM EVALUATION, TEST CONSTRUCTION, TESTING PROGRAMS, TESTS
IDENTIFIERS INDIVIDUALLY PRESCRIBED INSTRUCTION, IPI

ABSTRACT

IN REVIEWING THE INDIVIDUALLY PRESCRIBED INSTRUCTION (IPI) TESTING PROCEDURES, THE AUTHOR NOTES THAT THE ASSESSMENT OF STUDENT ACHIEVEMENT AND THE PLANNING OF INDIVIDUALIZED INSTRUCTIONAL SEQUENCES INVOLVE VALID AND RELIABLE MEASUREMENT OF MANY OF THE RELEVANT CHARACTERISTICS OF THE LEARNER. HE CRITICIZES THE PLACEMENT TESTING PROCEDURE AS INADEQUATE, AS IT DOES NOT ALLOW FOR THE EMPLOYMENT OF A GENERAL PSYCHOMETRIC MODEL. THE UNIT PRETESTS, WHICH FUNCTION AS DIAGNOSTIC TESTS, PROVIDE ONLY AN OUTLINE OF THE CURRENT STATE OF THE STUDENT'S KNOWLEDGE IN A SUBJECT, BUT DO NOT INCLUDE SUFFICIENT ASSESSMENT OF THAT INDIVIDUAL'S LEARNING HISTORY TO MAKE AN ADEQUATE PRESCRIPTION. THE CURRICULUM-EMBEDDED TESTS SEEM TO FUNCTION ADEQUATELY. THE UNIT POSTTESTS PROVIDE INFORMATION ABOUT A STUDENT'S MASTERY OF INDIVIDUAL ITEMS, BUT NOT ABOUT HIS ABILITY TO COMBINE THESE SKILLS TO A MORE GENERAL APPLICATION. IMPROVEMENT OF THE TESTING SYSTEM DEMANDS CAREFUL EXAMINATION OF GOALS AT ALL LEVELS OF ANALYSIS AND, IN PARTICULAR, DEMANDS THAT ALL ELEMENTS UNDER EXAMINATION BE RELATED TO THE ULTIMATE PURPOSES OF THE SYSTEM. (JY)

Draft

9/12/68

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED036185

Nitko, Anthony J.

MEASUREMENT OF INSTRUCTIONAL OUTCOME

VS

MEASUREMENT FOR INSTRUCTION:

A VIEW OF IPI TESTING PROCEDURES

September 1968

EM007771

Measurement of Instructional Outcome vs. Measurement for Instruction: A View of IPI Testing Procedures

Introduction

The formative evaluation of an instructional system involves the examination of its current operations in an attempt to answer the question of how the system can be improved. In the Individually Prescribed Instruction (IPI) project a number of operations can be examined. This paper examines one of these operations--the IPI testing procedures.

Essentially, the present use of tests developed for IPI are seen to conform rather closely to an experimental learning paradigm in which there is a test, a treatment, and a re-test. While this design is appropriate for measuring the outcomes *of* instruction, there is still the unanswered question of measurement *for* learning. The measurement of learning outcomes is no doubt a *necessary* condition for subsequent instructional planning. In an instructional system which achieves complete individualization, however, measurement of learned behavior is not a *sufficient* condition for the accomplishment of this goal. The discussion which follows is designed to present a rationale which might be termed *measurement for instruction*.

The first section of this paper is directed toward providing a rationale or some "lines alone which to think" concerning the currently existing IPI testing procedures. The second section will provide a proposal for an additional test battery which will function in a manner similar to the currently published and commercially available achievement tests. While similar in function to such tests as ITBS, it would not be identical and would be oriented toward a "content-referenced" pole rather than a "norm-referenced" pole.

I. IPI Testing Procedures

In the IPI instructional procedure the role of student assessment via formal testing procedures is of utmost importance. In fact, the very basis for the individualization of instruction is the assumption that individual capabilities can be validly and reliably determined (i.e. measured).

As a consequence, an elaborate test-oriented model can be formulated*

---

*Cox and Lindvall

---

which precisely describes the procedure through which an individual pupil can effectively attain mastery of all of the skills delineated by the three principle IPI curricula: Mathematics, Reading, and Science. This instructional procedure can be illustrated by a flow chart as in Figure 1. From this figure four major testing elements can be determined: Placement testing, unit pretesting, unit monitor testing (CET), and unit posttesting.

It is the purpose of this first section to examine each of these testing elements to determine their present operational functions and to examine how well each of these meets the theoretical model* from which they were originally established.

---

*Glaser

A. Placement Testing

In order to begin an instructional sequence which is individualized, IPI makes the assumption that every pupil possessed a different degree of mastery of the curriculum in which he is to study. As a consequence,

this degree of attainment within a given curricular area must be assessed for each and every pupil. The instructional procedure is so established that the initial entry point into a planned sequence of instructional units must be determined before any instruction can begin.

On closer examination, it is seen that the determination of this initial entry point into the curricular sequence is the most important consideration of placement testing. Once this point of entry is determined (assuming that it is the "proper" or "best" entry point for an individual student) effective instruction can begin. If the instructional units are sequenced in prerequisite hierarchies, assessment of the degree to which the entire curriculum (or a relatively large segment of it) is mastered is immediately known by an examination of the position of the predicted unit in the curricular sequence. Formal diagnositc testing of curricular mastery over a wide range of units would seem to be an inefficient means of initiating an instructional sequence.

If curricular units are not sequenced in prerequisite hierarchies, the problem of placement testing becomes much more complex. The prediction of the "best unit" in which to <u>begin</u> instruction is still the first question to be answered. However, the answer to the second question: "Where does instruction continue after this 'best unit' is mastered?", remains unanswered. Perhaps in this instance the curriculum itself ought to be restructured in order to provide for more effective individualization.

Beginning with this assumption--i.e. placement tests functioning as predictors--may or may not lead to the same type of tests which are currently employed with IPI. The assumption, in effect, allows for employment of a mor general psychometric (and learning) model. One basic deduction from this assumption, is that the placement tests can include <u>any corrolates</u> which make for effective and efficient predictors--whether or

not these corrolates have any "face validity" in the achievement (skills) domain. Restriction of the placement tests to only curricular stimuli (skills), limits the learning model by excluding other learning variables such as IQ, pupil motivation, need for review, and possible mediating responses. This restriction also excludes the possibility that skills not "mastered" on the placement test could be learned or relearned by doing work in another (perhaps higher level) unit of instruction. Measures of rate of learning then becomes measure of speed through a unit of instruction which is not necessarily learning rate since rate of learning tends to presupposes content naivete before instruction.

It seems logical, then, that placement testing ought to be solely predictor testing. At present, these tests attempt to perform two functions: predicting and diagnosing (i.e. measuring achievement or mastery). There is a growing body of evidence that neither of these functions is being adequately realized.

B. Unit Pretests

In the current IPI testing (and instructional) procedure, unit pretests are administered to every pupil on the basis of the results of his placement test scores. These tests are "in depth" tests of the skills which comprise the unit for which a preliminary assignment has been made. They test all skills which are indicated as "belonging" to a particular unit. The posttest for a unit is an equivalent form of the pretest. The information currently obtained from these tests is the identification of the skills within a unit which are mastered or are not mastered. At present these tests function (and indeed ought to function) as diagnostic tests.

The question which arises is whether this diagnosis is complete. Identifying mastered and unmastered skills within a unit is analogous to a physician's cataloging of a patient's symtoms. While knowledge of symtoms is a necessary condition for prescription development it is not a

sufficient condition. In the case of the physician some sort of additional information must be obtained (perhaps, patient history and other non-symtom characteristics) and this together with symtom information is processed by the physician before a prescription is developed. Sometimes additional tests need to be taken. At other times a case history needs to be obtained (eg. existence of alergic reactions which are non-symtomatic to the immediately prevailing condition but which could effect the patient if certain prescriptions were given to him). Finally, the cause of the symtoms is determined and an effective prescription ("cure") is developed.

The point here is that currently the IPI procedure does not provide systematic and reliable data as to the "cause" (and hence suggestions for the "cure") of skill deficiency. Prescriptions are written on the basis of "symtoms" and appear to be principally a function of the teacher* rather

---

*Swanson

---

than the individual student.

A re-examination of the pretest may reveal that "pretest" ought to be interpreted more broadly to formally include such things as: pupil learning history, pupil aptitudes for long range learning in the topic under study, skill mastery of the unit under consideration, and specific learning variables which are related to the particular skill which is to be learned. In this way an "individualized" prescription can be developed. For example, a step-by-step learning procedure may be appropriate for a relatively slow learner but not for a "fast" learner. Again, a different prescription would be written for a student who does not pronounce the "a" in a particular word and a student who mispronounces an "e" in the same

word. Knowing that the student cannot pronounce the particular word (i.e. has not mastered the skill) is not sufficient to write an effective prescription. Another example is the situation in which a given skill is better mastered by some students with paper and pencil material and the same skill is better mastered by other students through the use of some manipulative device.

The current assumption of IPI is that the teacher (like the physician) is taking all factors into account. There is no evidence that this is indeed occuring. In fact, the contrary is probably true. The reason for this may be in the pretesting procedures (broadly interpreted, again) which do not allow for accessible and reliable information to be provided to the teacher so that she may make a proper and meaningful differential diagnosis. Simply stated, the pretesting procedure ought to be one which has implications for instructional treatment on an individualized basis.

C. Curriculum Embedded Tests

Curriculum embedded tests function as criterion tests of within unit skills. As such they monitor the progress of each individual pupil. The CET's function in the same manner as teacher-made tests in the typical traditional instructional program with the exception that they monitor pupil progress more frequently and over a shorter (or more specific) learning experience. They appear to be adequately serving their function at the present time.

D. Unit Posttests

As currently used, unit posttests are "equivalent" forms (logical or content equivalence has been adhered to during their construction, but empirical equivalence has not been demonstrated) of the unit pretests.

That this practice of equivalency is open to discussion has been indicated above in connection with the unit pretests. As currently employed, unit posttests attempt to assess mastery of every skill within the unit in the same manner as does a unit pretest.*

---

*Cox and Boston

A closer examination of the function of these unit posttests seems to reveal a basic inadequacy in their stated purpose. To more clearly understand the purpose and function of a unit posttest one must consider the function of the unit itself.

Examination of the various IPI curricula and discussion of these curricula with specialists who are concerned with them reveals that the skills within a unit of instruction take one or both of two forms: they are sequenced or practically sequenced and lead to a terminal skill or goal for the unit; or they are not sequenced and lead to one or many terminal behaviors for the particular unit. What is often sought as a result of the instruction within a unit is to have a student exhibit all or many of the skills in combination (terminal unit behavior). Very often the "skills" which are catalogued within a unit are not instructional objectives, but observable pupil activities which occur in the course of instruction (these are often termed "teaching objectives" or "teacher techniques"). On closer examination one discovers that what is really sought as the terminal behavior is often some "process behavior"* and mastery of the within unit (specific)

*Glaser

content-oriented skills is of no concern. Very often what is sought is the student's ability to generalize or apply the specific content within the unit to a new and previously unexperienced situation (at the appropriate level of difficulty, of course).

Viewing the unit posttest as functioning as only a skill within unit test does not provide for the above (more meaningful) measurement. As an example of what might happen when the skill-by-skill approach is taken consider a situation in which five behaviors (skills) are sequenced within a unit ($R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_4 \rightarrow R_5$). Suppose further that each response builds upon the other and, in particular, R_5 involves using all previous responses in a new situation (i.e. other than the one employed in the instructional unit itself). As an additional condition assume that a perfectly reliable and valid measure [test] has been built which assesses the mastery of each of these behaviors. The instrument is such that it gives equal weight to each behavior (both logically and in the contribution-of-variance sense) and a mastery criterion of 80% correct is established. It is easy to see what may happen: A student demonstrates the mastery criterion of 80% by correctly making the first four responses. The fifth, and most important response to be learned in the unit, has not been mastered and the student is allowed to proceed to higher-order units which presuppose that this important behavior is in his repertoire.

E. Summary

The above sections attempted to review the IPI testing procedures in an effort to integrate these procedures into a meaningful model of individualized instruction. The basic point of departure in these discussions is the conviction that assessment of student achievement and the planning of individualized instructional sequences involves valid and reliable measurement of many of the relevant characteristics of the learner. This measurement is seen as part of the instructional program, and not simply something which intervenes in the instructional process from time to time and which has limited value. Perhaps the IPI measurement procedures function analogously to a governor on an engine, which allows engine progress or impediment according to some pre-established principal. This is a decidedly different situation from a roadside radar device, for example, which would simply measure engine progress and report that progress. But even this analogy is inadequate.

One essential aspect of the view of IPI testing presented in the foregoing discussion is that these tests function to provide the teacher and the learner with information concerning what instructional procedures ought to be immediately undertaken by the learner. This function is radically different from the traditional measurement conception which posits that achievement testing involves the assignment of a metric to passed experience or what has been already learned. Although assessment of behavioral repertoires already obtained is a necessary purpose for testing in IPI it is not a sufficient purpose. That is, IPI testing ought to be interpreted broadly as IPI measurement which has implications for immediate teaching and learning. This conceptualization allows for a much broader base of operation than does the usually psychometric view which states,

in effect, "Name a behavior (precisely, of course) and I'll measure." Tests are instruments which have very specific functions and allow certain kinds of decisions to be made when the "reading" of this instrument is taken. They are more than collections of somewhat related items chosen for administrative convenience or on the basis of arm chair philosophical postulates. This is analogous to the position that a curriculum is more than a collection of statements of behavioral objectives which seem to be related or which are grouped for convenient teaching practice. Just as curricular objectives need to be made meaningful to the entire curriculum, so too, curricular tests need to be made meaningful to the entire curriculum and instruction process.

The conception of the role of IPI testing which has been described above requires a scrutiny of the instruments currently functioning in the IPI project. Basic psychometric considerations of what is meant by test reliability and validity in the IPI context need to be undertaken. Are traditional approaches to these measurement problems appropriate? If so, under what conditions? If not, why not? But what is needed is more than theoretical debate. The hypothesis generated need to be empirically examined. New hypothesis can then be generated and tested.

A most basic consideration for any test development must begin with a clear delineation of the behaviors which need to be measured. Since worthwhile products take time to be developed and tested, a production schedule with frequent changes in the definition of the product (i.e. intended behavioral outcomes of instruction) mitigate against the development of these instruments as worthwhile. In some cases it may be necessary to re-examine and reformulate an entire set of curricular specifications

before any attempt at test construction can begin. One cannot combine hastily prepared curricular objectives with hastily prepared measurement instruments and then "field test" the entire conglomeration with the purpose in mind of drawing a conclusion concerning the effectiveness of IPI as an instructional system. IPI cannot be improved by cursory piecemeal examination. Improvement of a system demands careful examination of goals at all levels of analysis and, in particular, demands that all elements under examination be related to the ultimate purposes of the system.

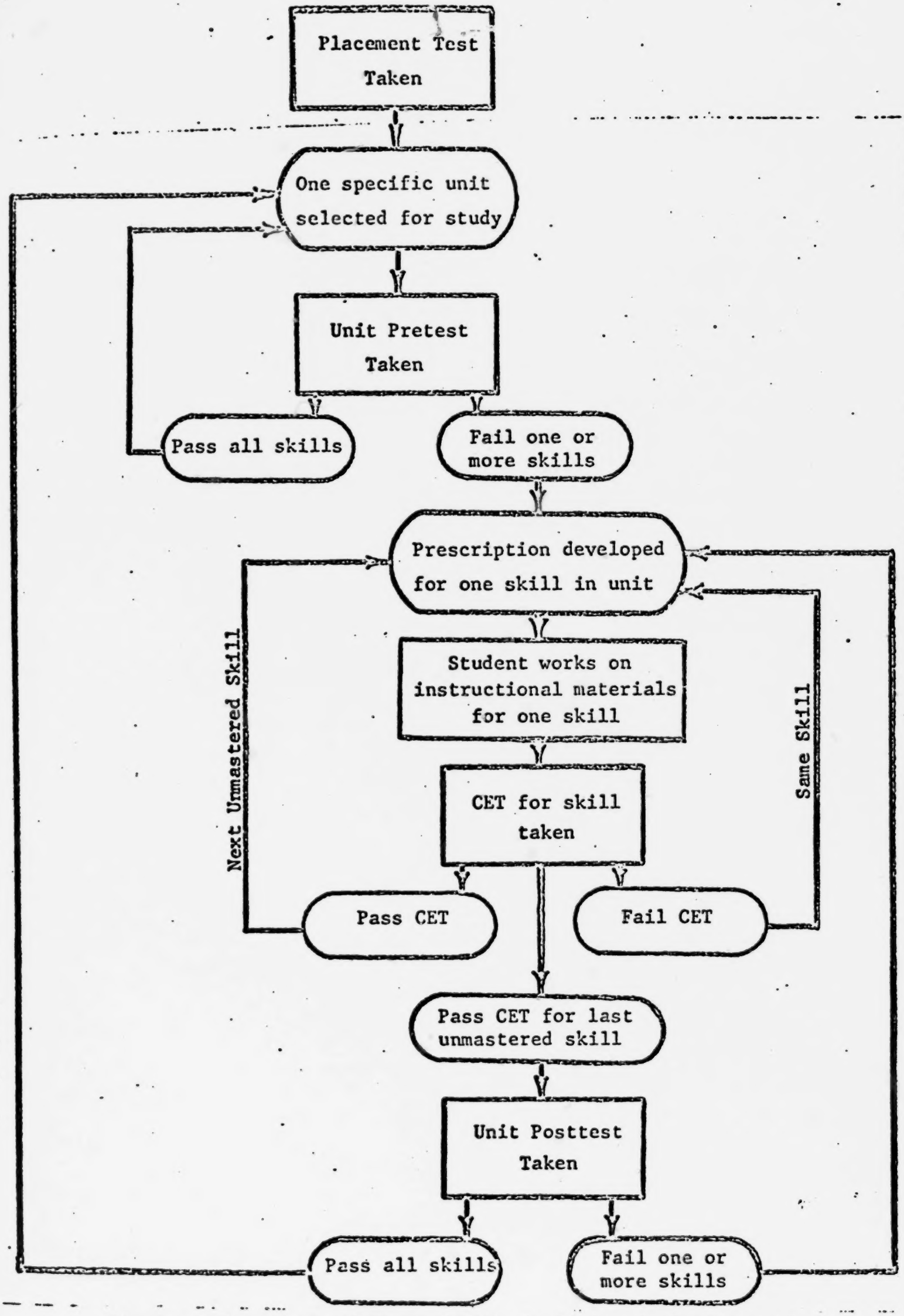
Placement Test Taken
One specific unit selected for study
Unit Pretest Taken
Pass all skills
Fail one or more skills
Prescription developed for one skill in unit
Student works on instructional materials for one skill
CET for skill taken
Pass CET
Fail CET
Next Unmastered Skill
Same Skill
Pass CET for last unmastered skill
Unit Posttest Taken
Pass all skills
Fail one or more skills

Figure Flow-chart of Steps in the Cycle for Evaluating and Monitoring of Pupil Progress in the IPI Procedure